

MENTAL TASK CLASSIFICATION BASED ON ENTROPY, SPECTRAL ENTROPY AND MUTUAL INFORMATION

F. Mamashli¹, A. M. Nasrabadi², SH. Shobeih

¹ Department of Biomedical Engineering, Poly Technique University, Tehran, Iran

² Department of Biomedical Engineering, Shahed University, Tehran, Iran

f_mamashli@yahoo.com

Abstract—with recent advances in signal processing and biomedical instrumentation, EEG signals can be used as a new communication channel between human and computer. Implementation of this channel is possible through recording and analyzing brain waves. Such a system translates human thoughts for a computer thus it is called a brain computer interface (BCI). There are lots of ways for feature extraction. Shannon's communication's theory may be used to quantify the information of BCI data. In this paper, we aim to evaluate the performances of different features extracted from the EEG signal which suits best for this purpose. We use Fourier coefficients, entropy, spectral entropy and mutual information as features to a linear discriminant analysis (LDA) classifier for the classification of 3 different mental tasks mainly: baseline, rotation and multiplication. We compare their results on 4 different subjects taken from the Anderson's database. Results show that classification using combination of entropy, spectral entropy and mutual information as features outperform all the other features used for classification.

Keywords—Brain computer interface, Entropy, Spectral entropy, Mutual information, Linear classification

I. INTRODUCTION

Although the first research relating to BCI's appeared in the 1960's, it still in its infancy for a variety of historic reasons. Recent changes in technology and advances in research have changed the environment for BCI research. Recent advances include enabling owl monkeys to control a robotic arm and a computer cursor through their thought alone in 2000, and enabling a paralyzed 25-year old man to do the same in 2004.

The basic model of a BCI system is shown in figure 1. This diagram will be the basis of the rest of this section, describing the brain signals, signal acquisition, signal feature extraction, signal feature translation, and some possible outputs.

BCIs are capable of monitoring various brainwave phenomena. Examples of brainwave phenomena include slow cortical potentials (SCPs), P300 potentials (positive peaks after 300ms), and mu or beta rhythms recorded from the scalp. Methods to observe these phenomena include EEG, magnetoencephalography (MEG), positron emission topography (PET), functional magnetic resonance imaging (fMRI), and optional imaging. Other signal features that are considered to contaminating the user intent are electromyography

(MEG), and electrooculography (EOG), which result from muscle and eye movement respectively. There are also invasive BCI recording methods (e.g. implanted electrodes).

Most BCIs use EEG signals, which represent the electrical activity in the brain as measured from outside of the skull [1]. EEGs are normally not only providing insight concerning important characteristics of the brain activity but also yield clues regarding the underlying associated neural dynamics. The processing of information by the brain is reflected in dynamical changes in this electrical activity. The ensuing activity variations are found in (i) time, (ii) frequency, and (iii) space. The EEG-signal is what mathematicians call a non stationary time-series (ST). Powerful analytical methods have been developed over the years to extract information from ST [2]. They can be characterized as time-domain or frequency-domain (or both). This information should help the BCI system to distinguish the parts of the signal that encode the user's intent. In addition to spatial filtering, a variety of options for feature extraction are currently under study by others, including spatial and temporal filtering techniques, signal averaging, voltage amplitude measurements, and spectral analysis (e.g. by using the Fast Fourier transform)[1].

Entropy and spectral entropy has been mostly used as features in different fields such as Neonatal Seizure Detection and speech recognition [15]. It is a measure of disorganization or uncertainty in a random variable. The information can be interpreted as essentially the negative of the entropy, and the negative logarithm of its probability [10].

Feature extraction based on entropy and spectral entropy will improve accuracy in classification results due to feature reduction from 30 to 6 in Fourier coefficient rather than entropy and spectral entropy.

The commonly used EEG features for this purpose include the use of Fourier coefficients, entropy, spectral entropy (SE) and mutual information. Entropy is usually used in the context of pattern classification and information technology. Originally the entropy was defined for information sources by Shannon [9].

Mathematical Background

In 1948, Claude Shannon defined the information theory

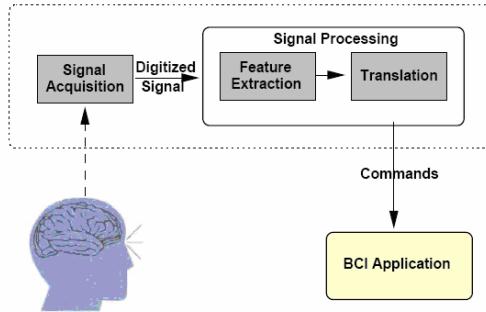


Fig. 1. BCI Procedure

concept of entropy which he described as the mean or the expected value of the information, where entropy measures the average uncertainty or the average self-information conveyed by an event [4]. The self-information represents the information that is conveyed by an event that occurs with probability $p(x)$. Since self-information measures uncertainty, events that are uncertain contain a lot of information, therefore, have higher entropy. Therefore, the higher the Entropy, the greater the uncertainty as to which event will occur. The mathematical representation of entropy is found as follows: If we let x be a discrete random variable, the entropy is denoted as:

$$H(x) = -\sum_{i=1}^n p(i) \log_2 p(i) \quad (1)$$

Where Ex is the expectation with respect to x , $i(x)$ is the self-information and $p(x)$ is the discrete probability distribution function (PDF). The discrete PDF is estimated by dividing x into N different amplitude bins and determining how many values of x are in each bin and normalizing by the number of values. Any logarithm base may be used for the equation as long as the base is maintained. In the equation above, “base 2” is used: therefore, the entropy is calculated in terms of bits. Shannon’s definition has been applied, modified, and proven valid in a variety of fields. In 1979, Powell and Percival introduced the concept of spectral entropy (SE), based on the peaks of the Fourier spectrum, as a measure of regularity [6]. Inouye claimed that the spectral entropy is a useful means for observing the degree of EEG irregularity [7]. To find the SE, first the discrete-time form of the spectrum is found from:

$$p(k) = \frac{1}{NT} |TX(k)|^2 \quad (2)$$

$$X(k) = \sum_{n=0}^{N-1} x[n] \exp(-j2\pi k/N) \quad (3)$$

Where $X(k)$ is the discrete Fourier transform (PDF), T is the sampling interval and N is the number of samples in the data set. The SE then is expressed as [3]:

$$H = -\sum_k P(k) \log_2 P(k)$$

The third feature in our study was mutual information which has been widely used in feature selection and Alzheimer’s disease study. The expected mutual information is a measure of dependence between two dimensions, i.e., to what extent events tend to co-occur in particular combinations. In this respect it is comparable with the product-moment correlation coefficient in the way entropy is comparable to the variance. Mutual information is given by:

$$J(X, Y) = \sum_{i=1}^m \sum_{j=1}^n p_{ij} \log_2 \left(\frac{p_{ij}}{p_{i \cdot} \cdot p_j} \right)$$

Sometimes also denoted by $M(X, Y)$ or $T(X, Y)$. It can be shown that $J(X, Y) \geq 0$ and that $J(X, Y) = H(Y) - H_x(Y)$ and $J(X, Y) = H(X) - H_y(X)$ (Theil 1972: 125-131). It can further be derived that the multi-dimensional entropy equals the sum of marginal entropies minus the mutual information (Theil 1972: 126):

$$H(X, Y) = H(X) + H(Y) - J(X, Y)$$

[16].

II. METHOD

A. The Data Set

The data set used in this study was obtained from the Colorado University. The data have been recorded according to the 10-20 standards from the C3, C4, P3, P4, O1, O2 electrodes. Each recorded signal has a length of 10 seconds with a sampling rate of 250 Hz. Seven subjects participated in the recordings, but we only used the signals of the first, third, fifth and the sixth subjects because of their larger number of trials [17]. The mental tasks considered in the data collection includes: baseline, mental multiplication and mental 3-D geometric rotation. Our goal is to classify mental tasks from each other using mentioned features and compare their results. The EOG signal which was recorded simultaneous to the EEG signals served to remove artifacts caused by eye movements.

B. Data Processing

Initially, using the EOG signals and an appropriate empirical threshold value, sections of the EEG signals coincident with the eye movements are identified and removed from the data set which leads to signals free of artifacts from eye movements. The resulting signals were then divided into windows of 2 seconds length each with an overlap of 1 second.

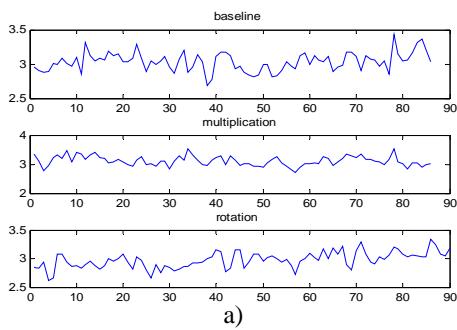
C. Extracting Features

Feature extraction is a process focused on discovering a pattern that can differentiate various classes. Usually

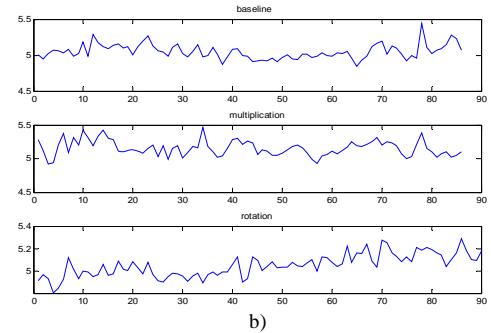
two approaches are used to extract the features from EEG signals. The first approach is based on the characteristic P300 signal following the occurrence of an event, referred to as event related potentials (ERP). Since this approach relies on a device that interacts between a subject and the stimuli corresponding to the event. It is impractical in our study. Hence, we employ the second approach [8]:

As previously mentioned we used several methods for extracting useful features from the EEG signal. They are described as follows:

- 1) *Spectrum analysis*: the power spectral density of signals is used as features in this application [8]. The PSD of the clean EEG signals is integrated over 5 frequency bands: 0-3.5Hz (delta), 4-7Hz (theta), 8-13Hz (alpha), 14-34Hz (beta) and Gamma > 35Hz [11]. Previous studies show that these frequency bands change characteristics while performing mental tasks [12], [13] and one of the best ways of detecting these changes is using their power spectral density [14].
- 2) *Entropy (E)*: As previously discussed Entropy is a measure of the amount of disorder in a signal. Another way of looking at is that entropy is how evenly spread a set of numbers is. An example of the entropy calculated for a trial of C3 EEG channel in three tasks is shown in fig. 2 a). This figure indicates that entropy level of EEG signals changes in different tasks but it has low variation in different channels. It means that the mean value of entropy is almost constant in 6 channels.
- 3) *Spectral entropy (SE)*: The feature was computed over each sliding window segment using the mathematical technique described in Section 1 with 125 point (1 second) of overlap. In our algorithm, we calculated the FFT as a fast means of obtaining the DFT. The SE was computed for all channels over each segment [3]. An example of the spectral entropy is shown in fig. 2 b).
- 4) *Mutual information (MI)*: Mutual information is widely used, in a descriptive way, to measure the stochastic dependence of categorical random variables. In our study we have estimated the mutual information between each segmentation of one EEG channel with the whole segmentation of the other channels. Therefore the corresponding feature vector consist 15 elements.



a)



b)

Fig. 2. EEG pattern in three different tasks:

a) Entropy b) Spectral entropy

D. Classification

The extracted feature were fed into a linear discriminant analysis classifier (LDA) using a probability model describing the discriminant function and supposing class distribution of $f_i(x)$ and prior probability of π_i . Where x is the observation of dimension q:

$$P(\text{class} = i | x) = \frac{f_i(x)\pi_i}{\sum_j f_j(x)\pi_j}$$

It can be shown that the rule that maximizes the conditional probability shown above will give the smallest number of misclassifications which is known as Bayes' rule. If we further assume that the classes have a Gaussian distribution with mean μ_i and covariance Σ then maximizing the conditional probability is equivalent to finding the i th class which maximizes L_i .

$$L_i = x^T \Sigma^{-1} \mu_i - \mu_i^T \Sigma^{-1} \mu_i / 2 + \log \pi_i$$

We apply maximum likelihood estimation for calculating μ_i , we arrive at a linear discriminant analysis [18].

III. RESULTS

In the simulations three classes have been classified with a feature vector of 6 dimensions for entropy and spectral entropy, 15 for mutual information and 30 for frequency features. The classification results from the LDA classifier are shown in Table 1. As it can be observed from the fourth row of this table, when we combine all of the entropy, spectral entropy and mutual information features as the input to our classifier with a total dimension of 27, the classification results yield much better. The reason behind the poor classifications results from subjects 3 and 5 might be from not to have attention. Hence, their respective EEG signals differ substantially with those of subjects 1 and 6.

IV. CONCLUSION

The purpose of this study was to introduce new features based on entropy and evaluate their performance as features for a BCI. We evaluated the performances of different features, mainly: entropy, spectral entropy, and mutual information and Fourier coefficients as inputs to a LDA classifier for the classification of three mental tasks performed by 4 subjects. Although the best classification results were obtained by the combination of entropy, spectral entropy and mutual information but results based solely using entropy features were promising as well.

Table 1
Classification results with test data using different features

	Sub1	Sub3	Sub5	Sub6
Entropy (E)	90.48 \pm 12.5	56.58 \pm 39.1	62.38 \pm 27	79.18 \pm 30
Spectral Entropy(S)	94.09 \pm 9.7	62.14 \pm 29	65.79 \pm 25	80.72 \pm 24
Mutual Information	84.23 \pm 18.5	57.57 \pm 30	54.61 \pm 29	84.51 \pm 16
E+SE+MI	96.58 \pm 4.4	66.74 \pm 35	71.16 \pm 23	92.5 \pm 12
Fourier coeff.	95.62 \pm 4.9	61.85 \pm 35	66.18 \pm 29	91.52 \pm 8.8

Table 2
Classification results with test data Using different features

	Sub1	Sub3	Sub5	Sub6
Entropy	91.98 \pm .9	61.3 \pm 4.7	65.32 \pm 1.8	81.68 \pm 1.7
Spectral entropy	94.3 \pm .5	66.79 \pm 3.7	67.18 \pm 1.6	82.21 \pm 2.8
Mutual information	89.2 \pm 1.5	64.86 \pm 3.08	60.42 \pm 3.2	88.58 \pm 1.99
E+SE+MU	98.3 \pm .4	78.06 \pm 3.6	78.39 \pm 1.8	97.23 \pm .7
Fourier coeff	97.45 \pm .32	75.48 \pm 4.13	79.46 \pm 3.2	96.33 \pm .56

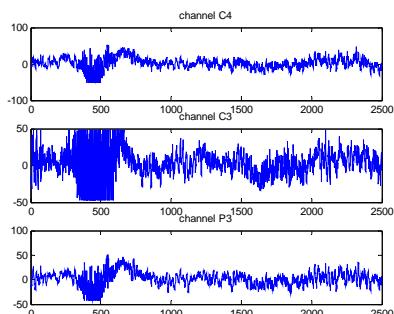


Fig 3. EEG signal from subject 3 in the second trial baseline

REFERENCES

[1] J. Thorpe, P.C. van Oorschot, Asomayaji "Pass-thoughts: Authenticating with our minds "Digital security Group, School of Computer, Carlton University, April 18, 2005.

[2] A. Plastino, O.A. Rosso "Entropy and statistical complexity in brain activity" *europhysics news Magazine*, pp. 224, 2005.

[3] M.M. D'Alessandro, G. J Vachtsevanos, R. Esteller, J. Echauz, A. Koblasz, B. Litt "Spectral Entropy and neural involvement in patients with Mesial temporal lobe epilepsy" *METMBS'2000*, Las Vegas, Nevada, June 26-29,2000.

[4] W. Gappmair, "Claude E. Shannon: The 50th Anniversary of information Theory," *IEEE Communications Magazine*, vol. 37, no. 4, pp. 102-105, 1999.

[5] C. E Shannon, Claude Elwood Shannon: collected papers. New York: IEEE Press, 1993.

[6] R. Q. Quiroga, Quantitative Analysis of EEG Signals Time Frequency Methods and Chaos Theory. Ph.D. Dissertation, Medical University of Lubeck, 1998.

[7] T. Inouye et al., "Quantification of EEG irregularity by use of the entropy of the power spectrum," *Electroencephalography*, vol. 79, pp.204-210, 1991.

[8] T. Lan, A. Adami, D. Erdogmus, M. Pavel "Estimating cognitive state using EEG signals" unpublished.

[9] C.E. Shannon, "A mathematical theory of communication," *Bell system Technical Journal*, vol. 27, pp. 379-423, 623-656, July, Oct. 1948.

[10] A. M. Toh, R. Tognoni, S. Nordholm, "Spectral entropy as speech features for speech recognition" unpublished.

[11] Ernst Niedermeyer, Fernando Lopes da Silva, "Electroencephalography: basic principles, clinical applications related fields", unpublished.

[12] Erol Basar, Canan Basar-Eroglu, Sirel Karakas, Martin Shurman, "Gamma, alpha, delta and theta oscillation govern cognitive processes", *International journal of psychophysiology* 39, 2001.

[13] G. Pfurtscheller, F. H. Lopes da Silva, "Event related EEG/MEG synchronization and desynchronization: basic principles", *Clinical neurophysiology* 110, 1999.

[14] S. Solhjoo, M. H. Moradi, "Mental task recognition: a comparison between some of classification methods", *BIOSIGNAL 2004*.

[15] S. Faul, G. Boylan, S. Connolly, W. Marnane, G. Lightbody, "A novel Automatic Neonatal Seizure Detection System" *ISSC, 2005*.

[16] K. Frenken "Entropy statistics and information theory" *The Elgar Companion to Neo- Schumpeterian Economics* July 2003.

[17] A. H. Omidvarnia, F. Atry, S. K. Setarehdan, B. N. Arabi "Kalman filter parameters as a new EEG feature vector for BCI Application" University of Tehran, Iran.

[18] F. Abdollahi, A. M. Nasrabadi, "Combination of Frequency Bands in EEG for Feature Reduction", *EMBC IEEE*, New Yourk, 2005.